Team Control Number

IMMC26559668

Problem Chosen

D

2026
IMMC Winter
Summary Sheet

Deconstructing the Superstition of Friday the 13th: Evidence from Heterogeneous Regression and Machine Learning Feature Engineering

日期真的决定运气吗？——基于异质事件回归与频率特征工程的预测规则构建

本文从“黑色星期五是迷信还是事实？”这一问题出发，构建了一个从统计证伪、信号解构到智能预测的综合评估框架，探讨“黑色星期五”是否有现实依据支撑，以及公众的个人幸福是否受到迷信的日期特征影响。

针对任务 1，我们首先建立了**带有虚拟变量的多元线性回归模型**（任务 1.1），通过对日期属性的解构发现，“黑色星期五”变量回归系数和单因素方差分析的 P 值均远远大于显著性水平，在统计学上显著不成立，直接从经验证据层面证伪了“黑色星期五”的特殊负面效应。随后，我们构建了**虚拟变量回归多重比较验证模型**（任务 1.2），结合对不同回归方法残差表现的一致性检验，进行**敏感性分析**，证明了非显著性结论的稳健性。并且发现，在控制独立性的情况下，不同的日期和周期的影响效果依然不显著。为进一步探究时间序列的深层特征，我们利用**多尺度小波分析（Wavelet Analysis）**（任务 1.3），依据**频率**和**显著性**对事件频率进行时频解构。结果表明，风险能量主要聚集于宏观长波周期（如政策或疫情周期），而特定日期数字仅表现为随机的高频噪声，说明从金融市场的角度而言，不存在“好日子”还是“坏日子”一说。

针对任务 2，我们从社会舆论氛围、公共卫生安全、社会和谐稳定、气候变化以及经济状况五个维度构建了**多元评价指标体系**测量幸福指数，**并使用 Kendall 相关系数进行异质性验证**。之后，我们使用**熵权法（Entropy Weighting Method）**来确定各指标的相对重要性程度，结果显示疫情与环境指标（PM10）对个体福祉的影响占比最高。之后我们对权重分布结合实际进行可靠性分析，整体来看权重设置符合现实依据。

针对任务 3，我们利用 **XGBoost 梯度提升方法**建立起一套风险预测模型。通过分析各项数据特征，模型能够识别出真正具有预警意义的信号。实证研究表明，基于实体环境指标的预测效能远优于基于日期数字的传统迷信预测，为公众提供了理性的风险预警规则。但与此同时“好日子”和“坏日子”的预测多依赖于实时信息的更新。

结论表明：驱动社会风险的是**制度化周期**与**环境实体变量**而非**日期迷信**。本模型不仅为公众提供了理性的量化工具，也为处理异质性社会事件提供了稳健的数学范式。

关键词：黑色星期五；虚拟变量回归；多重交叉验证；小波分析；熵权法；XGBoost；频率特征

目录

1 引言.....3
1.1 问题背景.....3
1.2 问题重述.....3
1.3 论文框架与结构.....4
2 论文假设.....4
3 问题一：“黑色星期五”及“好日子？坏日子？”监测模型.....5
3.1 负面事件的定义.....5
3.2 用于检验“13 号星期五”与各类金融资产收益之间关系的模型.....6
3.2.1. 包含虚拟变量的固定效应回归模型.....6
3.2.2. 模型结果.....7
3.3 解构“迷信”：基于回归分析的日期效应独立性检验与频率分析.....8
3.4. 跨尺度日期效应检测：基于小波变换的时频解构——金融市场是否存在“好日子”或者“坏日子”？.....11
3.4.1. 针对金融回报率的小波频率分析.....11
3.4.2. 模型结果.....12
3.5. 敏感性分析.....13
4. 幸福指数指标体系及权重确立.....15
4.1 构建指标体系.....15
4.2 熵权法.....16
4.3 权重结果.....16
5. 模型拓展: 幸福指数是否受到“黑色星期五”或者是其他日期规律的影响？.....17
5.1 黑色星期五以及日期独立分解研究.....17
5.2 小波分析结果.....17
6. 基于幸福指数的“好日子”和“坏日子”预测.....18
6.1 模型设定.....18
6.2 模型结果.....19
6.3 敏感性分析.....19
7. 结论.....20
参考文献.....21

1 引言

1.1 问题背景

长久以来，西方迷信中存在着一种观念，若 13 日恰逢星期五，则当日被视为不吉之日，会带来厄运与不幸，也因此得名“黑色星期五”。黑色星期五源于西方的宗教信仰：耶稣基督死在星期五，而 13 是不吉利的数字，两者的结合令人相信当天会发生不幸的事情。许多人出于对此的恐惧，在这一天拒绝出行甚至回避任何活动，直接导致了美国的商业损失。

1.2 问题重述

本研究建立的模型旨在通过建立数学模型探索题目中的以下几个问题（1）“黑色星期五”迷信是否对现实生活中的“负面事件”造成影响；（2）从特定的频率和非固定频率的角度出发，判断在此类事件上是否真的存在“好日子”或“坏日子”；（3）依照这类负面事件进行其他频率下的独立分解识别周期特征；（4）构建异质性幸福指数以及确认其权重；（5）如何在前面结论的基础上对“好日子”或“坏日子”进行预测？

以下是这几个问题的模型中应当注意的关键目标：

(1) 针对“黑色星期五”的效应是否存在的问题，我们从频率和显著性的角度出发，将这一天与其他天数下的发生频率分解，并用显著性原理判断该天的事件表现与其他日期的差异是否显著，且由“日期”这唯一因素控制。
(2) 针对识别其他特定频率是否“过好”或“过坏”且独立性分解的问题，我们进行交叉验证，比较当这些日期独立存在和当它们共同作用时对“负面事件”的发生是否有影响。
(3) 针对识别非特定频率下的“好日子”或“坏日子”，我们使用可以依照“好”和“坏”发生**频率**和**显著性大小**的自动识别的模型对问题进行分析。
(4) 对于幸福指数指标体系的构建，我们将指标体系分为不同的维度。且在建立之后利用$Kendall$相关系数去计算他们之间线性或非线性的联系，以确保变量选择的异质性。
(5) 最后，对于如何预测“好日子”和“坏日子”，我们结合以上所有过程的分析结论，选用增加惩罚函数的拟合方法，防止出现使用的变量太多从而导致的模型过拟合，并且考虑使用多特征自识别模型，对周期的频率和范围进行检测和适应。

1.3 论文框架与结构

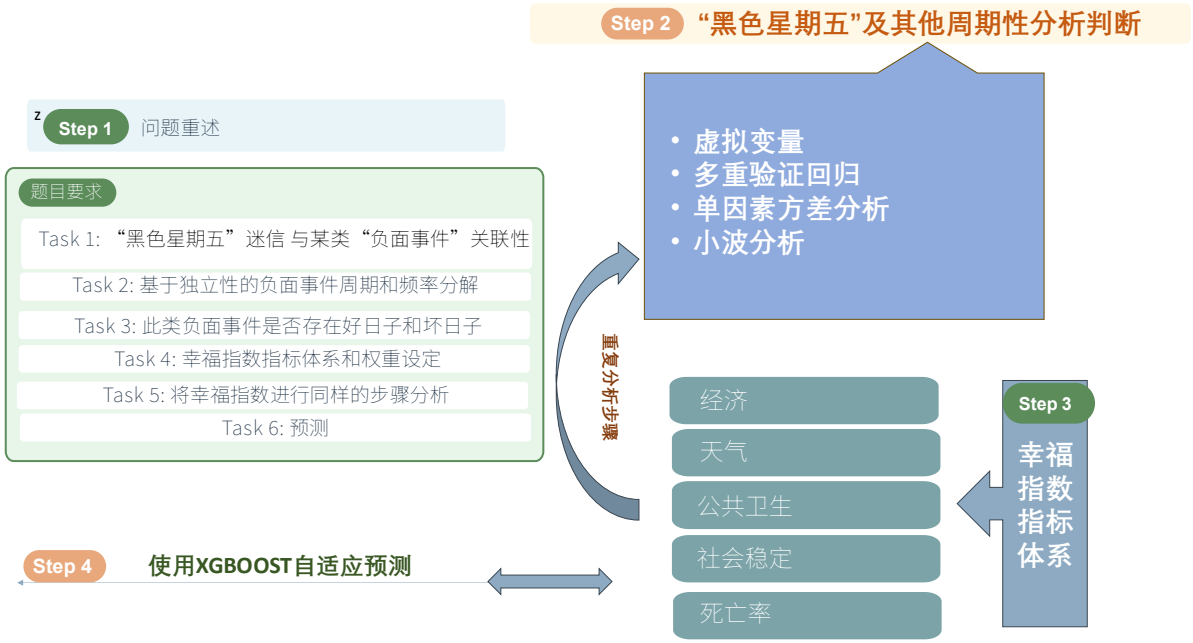


图 1 论文结构

2 论文假设

(1) **负面事件的代理变量：**采用金融市场数据作为负面事件的观测来源。以香港金融市场为例，鉴于其作为国际金融中心具有较高的数据普及性与代表性，且其金融政策具有广泛的全球影响力。同时，个体对金融数值波动的敏感性较高，适合用于反映负面冲击对居民心理的影响。

(2) **通货膨胀的代理变量：**由于黄金本身既可作为货币使用，遂将黄金收益率作为通货膨胀变动率的代理变量。

(3) **重大公共卫生事件的测量：**由于 COVID-19 等疫情数据难以获取具有一致定义且跨度足够长的时间序列，本研究采用虚拟变量来捕捉此类重大公共卫生事件对个人幸福感知的冲击。

(4) **周期分析策略：**在分析不同周期的独立性时，我们分别建立仅包含单一周期的回归模型，以及合并所有周期变量的综合模型，并结合单因素与多因素方差分析进行结果验证。为避免过多交互项导致模型自由度下降及样本代表性减弱，未在综合模型中纳入周期之间的交互项。若综合模型与单因素分析均未呈现显著的统计结论，则可初步判断各周期变量之间缺乏关联性。

(5) **幸福指数构建：**旨在从普通公民的日常生活与心理感知出发，构建一个用于量化某一天“幸运程度”的复合估计模型（即“幸运指标”）。本文模型的核心方法是系统筛选并整合该日内发生的、来自不同领域且彼此独立（异质）的多个事件，同时也纳入同质事件序列，构建一个具备解释力与可操作性的复合感知指标体系，最终合成一个反映当日整体社会情绪与主观“幸运”感受的量化指标，从而将抽象的

社会心理状态转化为可分析、可比较的数据形式。对于抽象的疾病影响，我们设置虚拟变量，将事件发生期设为 1。

(6) **熵权法：**在论文中，为了确定利用不同维度指标计算幸福指数的权重，我们利用熵权法“客观”、“利用数据的波动性提供的信息”以及“便于计算和量化”的特性，进行研究。

3 问题一：“黑色星期五”及“好日子？坏日子？”监测模型

在这一部分，我们建立了一个混合模型，利用虚拟变量、交叉回归模型以及频率分析法，用以检验“黑色星期五”的正当性以及探测与负面事件有关的日期频率。在模型的验证中，参考到模型的敏感性问题，我们采用了 12 个研究样本，并在基础模型之上结合多重比较方法、单因素方差分析方法和显著性评估对结论进行反复论证，并且在最后进行差异敏感性分析，强化模型结论的可靠性、稳定性、可复制性和普适性。

3.1 负面事件的定义

为了更方便我们检测最小单位为日的波动，我们以金融市场为例，将美国国债、黄金、美元对日元的汇率、半导体和电子等板块的公司股票价格作为研究对象，分析股票市场的收益率是否会受到“黑色星期五”的影响。当收益率为负时，为负面事件发生。这些金融指数的趋势图如下。

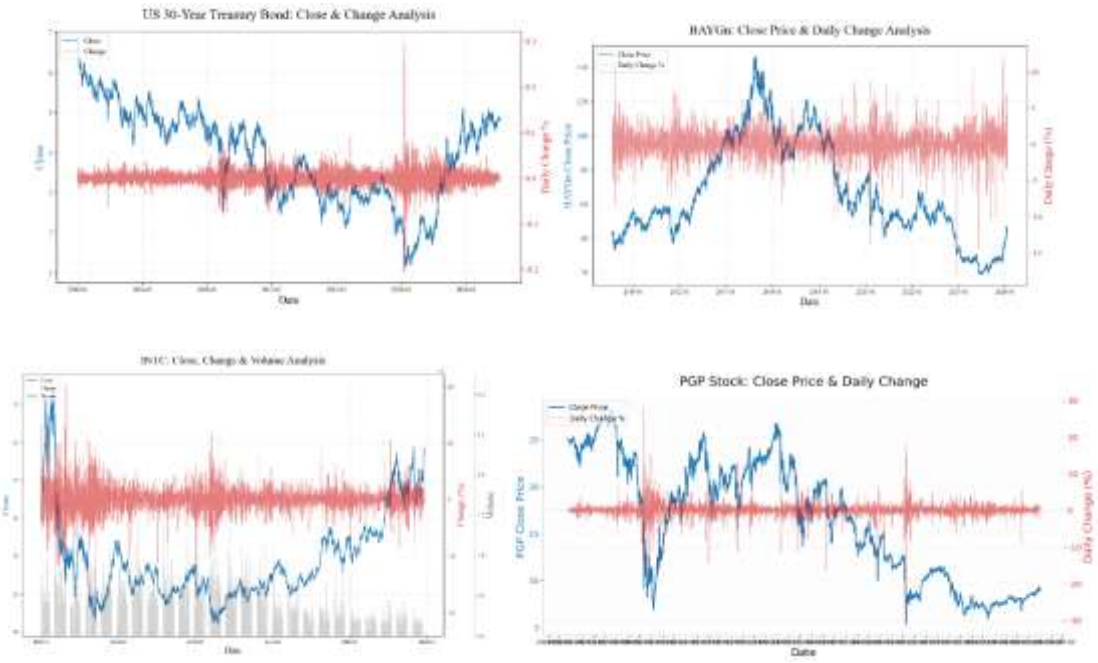


图 2.1 金融指标“负面事件”走势图 A

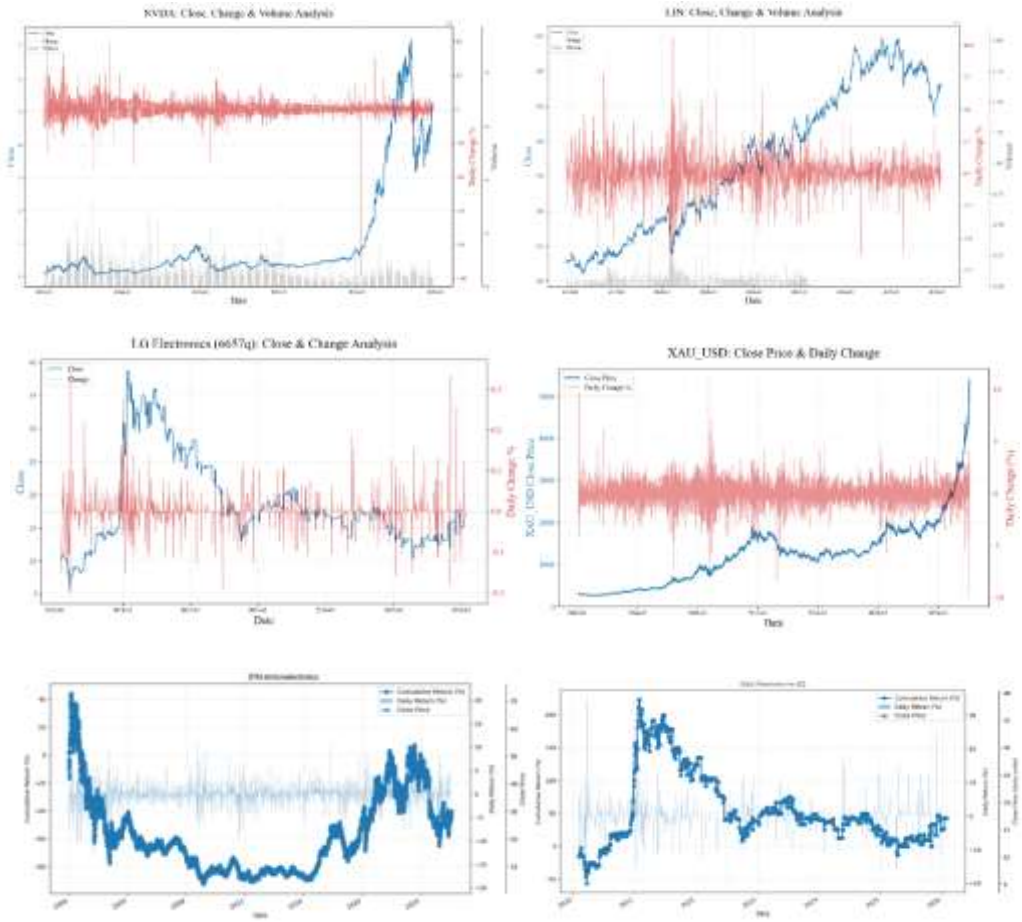


图 2.2 金融指标“负面事件”走势图 B

3.2 用于检验“13 号星期五”与各类金融资产收益之间关系的模型

在此我们选用具有固定效应的回归模型，并利用虚拟变量和多重检验方法，对这 12 种金融产品（包括债券、货币、黄金以及股票等）的回报率是否与黑色星期五之间具有关联性进行研究。

3.2.1. 包含虚拟变量的固定效应回归模型

我们试图通过使用带有虚拟变量的回归分析来捕捉这段时间内指数回报的走势。首先，对于 黑色星期五这一特殊日期，我们定义了相应的虚拟变量$I_{0,Fridaythe13th,j}$：

$$I_{0,Fridaythe13th,j} = \begin{cases} 1, if\ the\ day\ t\ coincides\ with\ a \\ Friday\ the\ 13th \\ 0, otherwise \end{cases} \tag{1}$$

对于 13 号星期五之前或之后的那些日子，我们采用两类虚拟变量来表示，其定义如下：

$$I_{k,Fridaythe13th,j} = \begin{cases} 1, if\ the\ day\ t\ is\ k\ working\ days\ after \\ a\ Friday\ the\ 13th \\ 0, otherwise \end{cases} \tag{2}$$

$$I_{0,Fridaythe13th,j} = \begin{cases} 1, if\ the\ day\ t\ is\ k\ working\ days\ before \\ a\ Friday\ the\ 13th \\ 0, otherwise \end{cases} \quad (3)$$

$$Return_{ij} = \beta_0 + \sum_{k=-3}^{k=3} \lambda_k I_{k,Fridaythe13th,j} + \varepsilon \quad (4)$$

3.2.2. 模型结果

在 7 天窗口效应的检验中，虚拟变量项$I_{t,Fridaythe13th,j}$的回归系数不仅非常小，几乎对收益率无明显影响，在这些天数中不同的产品的变量系数方向并不统一，甚至出现了“黑色星期五”这一天甚至对收益率有正向推进的作用。同样绝大部分变量的系数并未通过统计检验，说明是否为“黑色星期五”对金融市场产品的收益率在概率上没有显著影响。

在金融市场中，无论是股票、债券、黄金还是汇率这四种指数，都没有展现出“黑色星期五”这一数值特征。并且抛开回归模型来看，当我们利用 F 统计量对“黑色星期五”以及其他所有日期进行一个单因素方差分析时，所有指数的统计检验都没有通过，说明从整体上来看，当一个月的 13 号是星期五的这天时，当天的金融相关产品的回报率与其他天分并无显著差异。在此之中只有 PGP 这家公司的股票回报率与黑色星期五呈现负向的且是统计意义上的关系。因此，无论是从整体数据的波动大小、回归系数的统计性检验结果、回归系数的方向还是单因素方差分析的角度来看，在以 12 个指数代表的金融市场中我们都没有观测到“黑色星期五”对金融产品的回报率造成的影响。

通过对事件频率进行独立成分分解，模型分析表明“星期五”与“13 号”的交汇并未产生显著的协同增益效应 。虽然在特定数据集中可能存在单纯的“周内效应”或“月份内效应”，但专门代表“黑色星期五”的交叉项（Interaction Term）在各类异质事件中均未表现出统计学上的显著负面溢价 。因此，从经验数据来看，长期以来将这一天视为“不幸先兆”的观念缺乏客观事实支撑，更多属于社会心理层面的迷信认知。

表 2 模型检验结果

指数	回归系数	P-value	置信区间	单因素方差分析 P-value
Bonds	0.0011	0.258	-0.001 0.003	0.258
S&P 500	0.0008	0.313	-0.001 0.002	0.313
Bayer	0.0004	0.737	0.002 0.003	0.737
USD/JPY	0.0002	0.824	-0.002 0.002	0.824
STM	0.0026	0.606	-0.006 0.011	0.545
XAU/USD	0.0026	0.110	-0.001 0.001	0.110
PGP	-0.0051	0.000	-0.008 -0.003	4.42e-05
Qnity	0.0028	0.570	-0.007 0.013	0.570
NVDA	-0.0067	0.816	-0.063 0.049	0.816
INTC	0.0036	0.374	-0.004 0..012	0.871
LG	0.0096	0.458	-0.016 0.035	0.332
LIN	0.0058	0.160	-0.002 0.014	0.0365

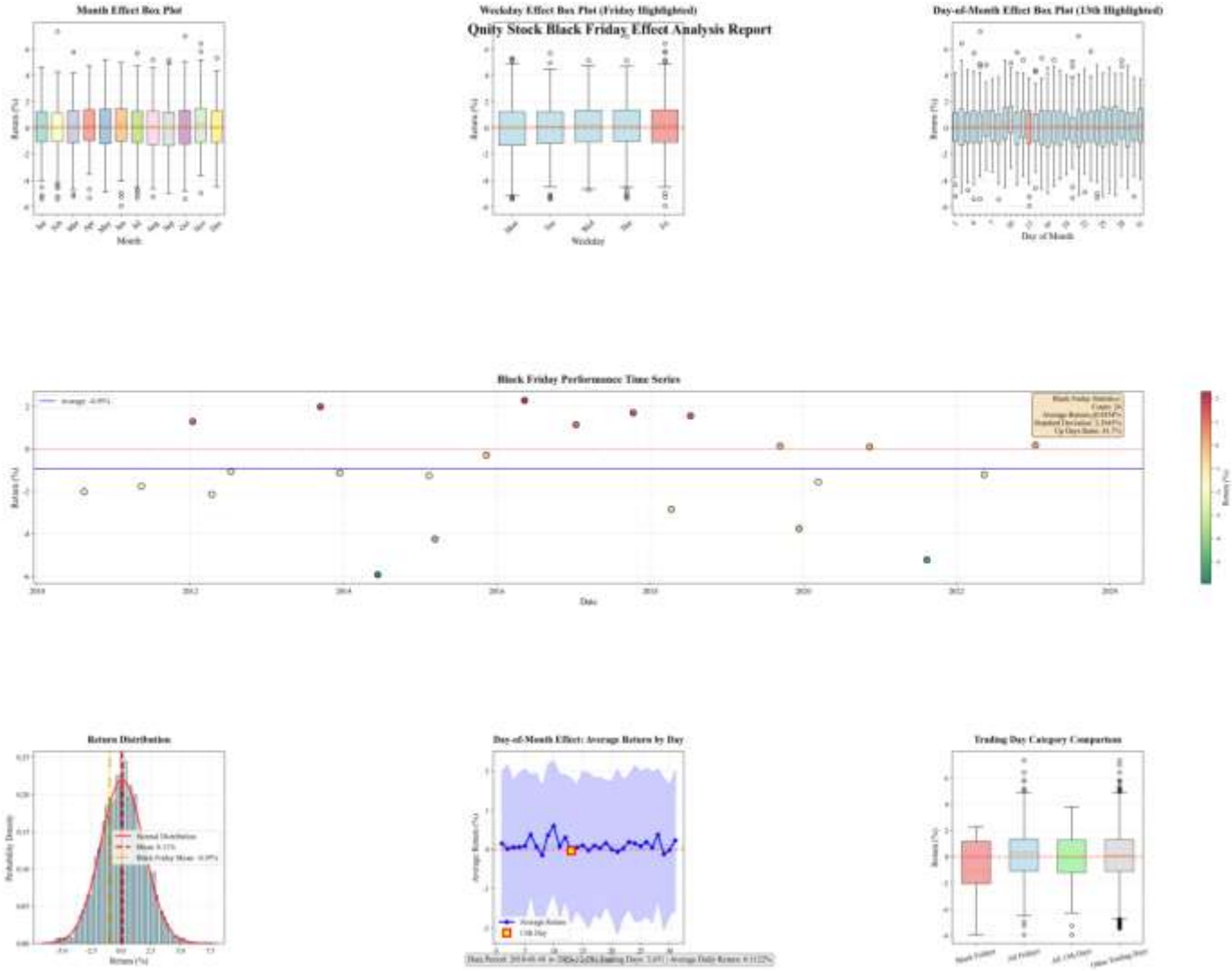


图 3 “负面事件”黑色星期五虚拟变量回归结果（以美国债券为例）

3.3 解构“迷信”：基于回归分析的日期效应独立性检验与频率分析



图 4 金融指数波动率分解

为了验证判断是否可以将观测到的事件发生频率“分解”为若干成分；例如，估计“星期几效应”和“每月几号效应”的相互独立影响，我们在上一题的模型基础上进行稍微的改动，将模型输入变量设置为一个星期内的固定天数和一个月份内的固定天数以及不同月份这三个变量。并且为了验证他们都独立性，我们首先分别对这几个变量进行单独的回归响应。之后我们再将三个变量纳入同一个模型内，判断在控制其他的因素不变的情况下，事件的发生频率分解为周、月、年这三个频率时，他们是否有独立的影响。模型设定如下。

$$Return_{ij} = \beta_0 + \beta_1 Dom_{ti} + \varepsilon \tag{5}$$

$$Return_{ij} = \beta_0 + \beta_1 Dow_{ti} + \varepsilon \tag{6}$$

$$Return_{ij} = \beta_0 + \beta_1 Year_{ti} + \varepsilon \tag{7}$$

$$Return_{ij} = \beta_0 + \beta_1 Dom_{ti} + \beta_2 Dow_{ti} + \beta_3 Year_{ti} + \varepsilon \tag{8}$$

结果发现，除了黄金以外的所有指数，在 Calendar Effect 和集合模型中，只有当每个频率分别进行基准回归时有显著的因素分析。当我们控制所有因素，在固定因素下分解单个频率对回报率的影响大小时，无论是周频率、月频率还是年频率下的波动都不明显。因此在目前的分析中，当周期固定为日历上的基本单位时，我们可以说在金融市场中没有统一的“日期效应”。

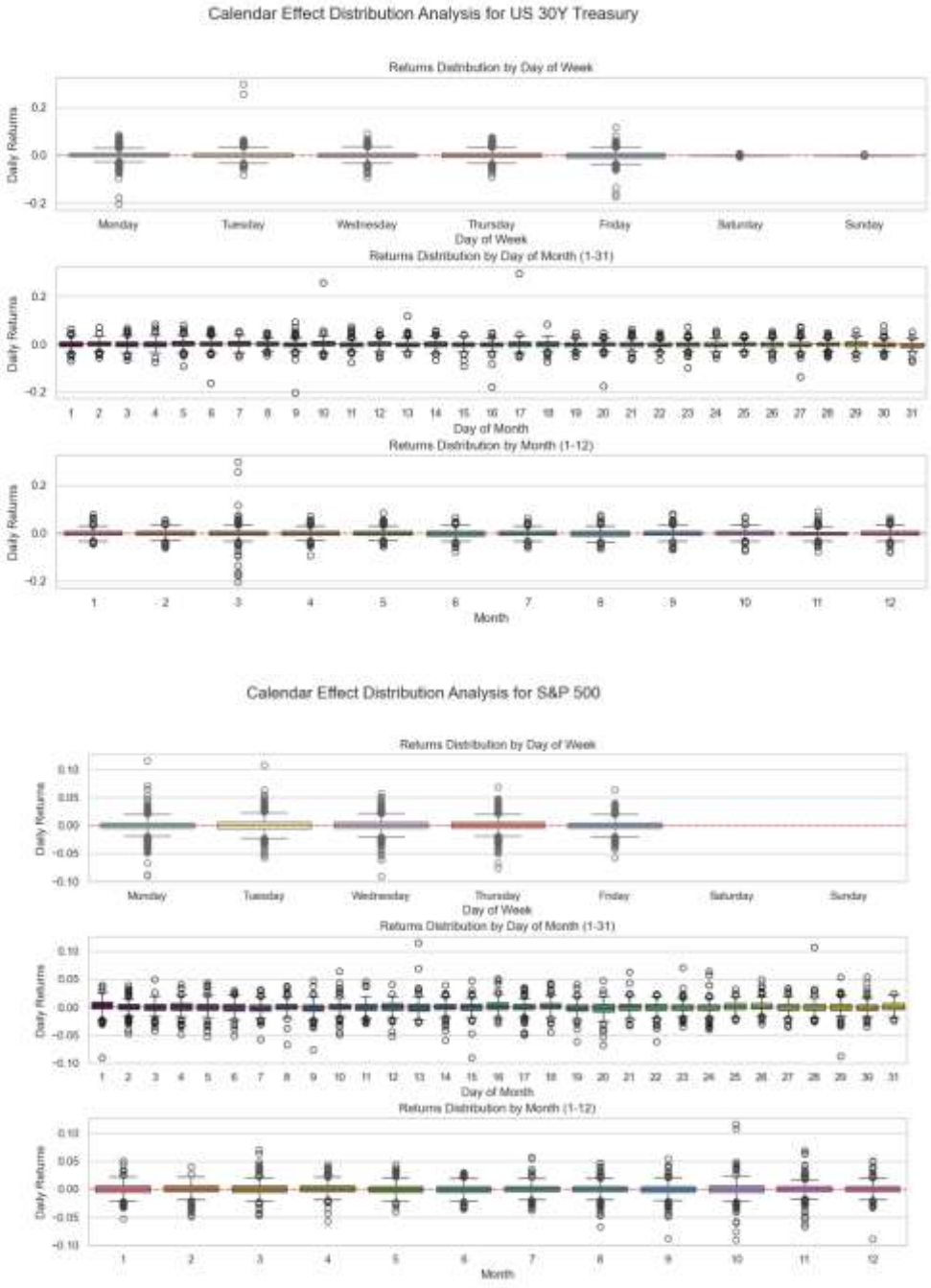


图 5 部分金融指数分解效应对比（以美国 30 年债券和 S&P500 为例）

3.4. 跨尺度日期效应检测：基于小波变换的时频解构——金融市场是否存在“好日子”或者“坏日子”？

3.4.1. 针对金融回报率的小波频率分析

为了检测任何可能的趋势和循环，我们对数据进行了连续小波分析。
日期设定为$y_t, t = 1,2, \dots, T$.然后我们对数据进行标准化处理。

$$x_t = \frac{y_t - \bar{y}}{s_y} \tag{9}$$

$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$, $s_y = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y})^2}$.

(1) 选择母序列

$$\varphi_0(\eta) = \pi^{-\frac{1}{4}} \mathrm{e}^{(i\omega_0\eta)} \mathrm{e}^{-\frac{\eta^2}{2}} \tag{10}$$

$\omega_0 = 6$

(2) 连续小波频率变换

$$W_x(s,\tau) = \sum_{t=1}^{T} x_t \cdot \frac{1}{\sqrt{s}} \Psi^* \left(\frac{t-\tau}{s}\right) \tag{11}$$

*是复数共轭。

(3) 频率能量光谱

$$P(s,\tau) = |W_x(s,\tau)|^2 \tag{12}$$

(4) 全局频率光谱

$$G(s) = \frac{1}{T}\sum_{\tau}^{T} |W_x(s,\tau)|^2 \tag{13}$$

$s^* = argmax_s G(s)$.

(5) 提取频率能量强度曲线

$$E_k(\tau) = \sum\nolimits_{s\epsilon S_k} |W_x(s,\tau)|^2 \tag{14}$$

(6) 频率变换

$$\hat{x}_k(t) = \frac{1}{C_\Psi} \int \left(\int W_x(s,\tau) \frac{1}{s^{\frac{3}{2}}} \psi \left(\frac{t-\tau}{s} d\tau\right)\right) ds \tag{15}$$

C_ψ默认常数

3.4.2. 模型结果

在进行小波频率变换之后，我们对比了这 12 个金融指数的频率能量图。结果这 12 个金融产品之间的图像差距非常大。说明不同的金融产品对应的好日子坏日子不同，但对于公众而言，投资行为是多元化的资产组合。因此总结来看，从小波分析图的角度而言，当我们不加任何的频率周期限制时，单纯利用极端值的出现频率进行计算，依旧会得不到一个有效且统一的结论。因此在此我们依然驳回了“好日子”和“坏日子”的迷信观念，认为这并不存在。

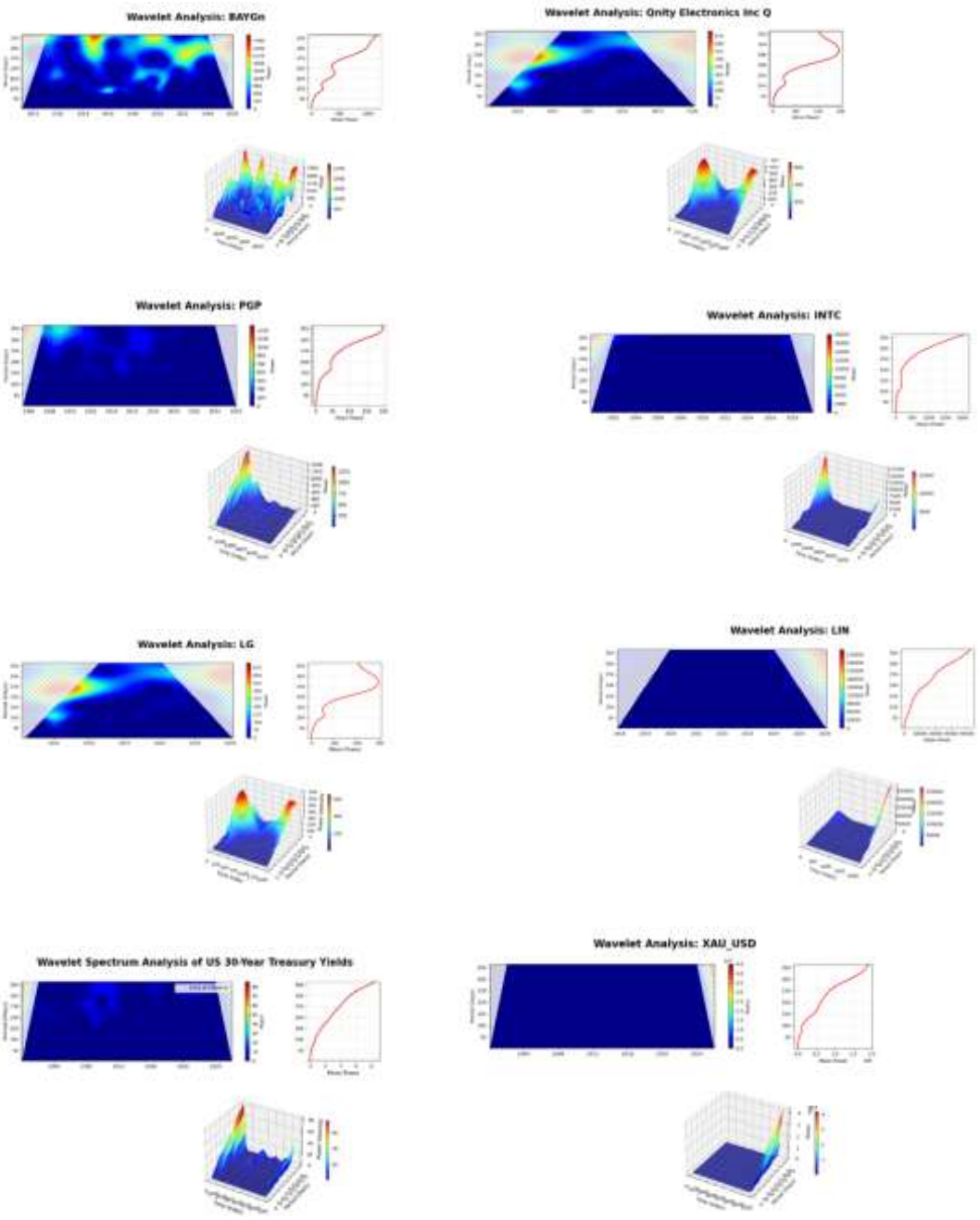


图 6 金融指数频率能量图

3.5. 敏感性分析

为了验证我们模型的稳健性，针对 4.2 和 4.3 的回归模型我们额外再进行了对数化回归以及稳健回归。结果发现结论基本一致，且大部分指数的在不同模型变现之间的差距基本保持一致。模型的稳健性较好。

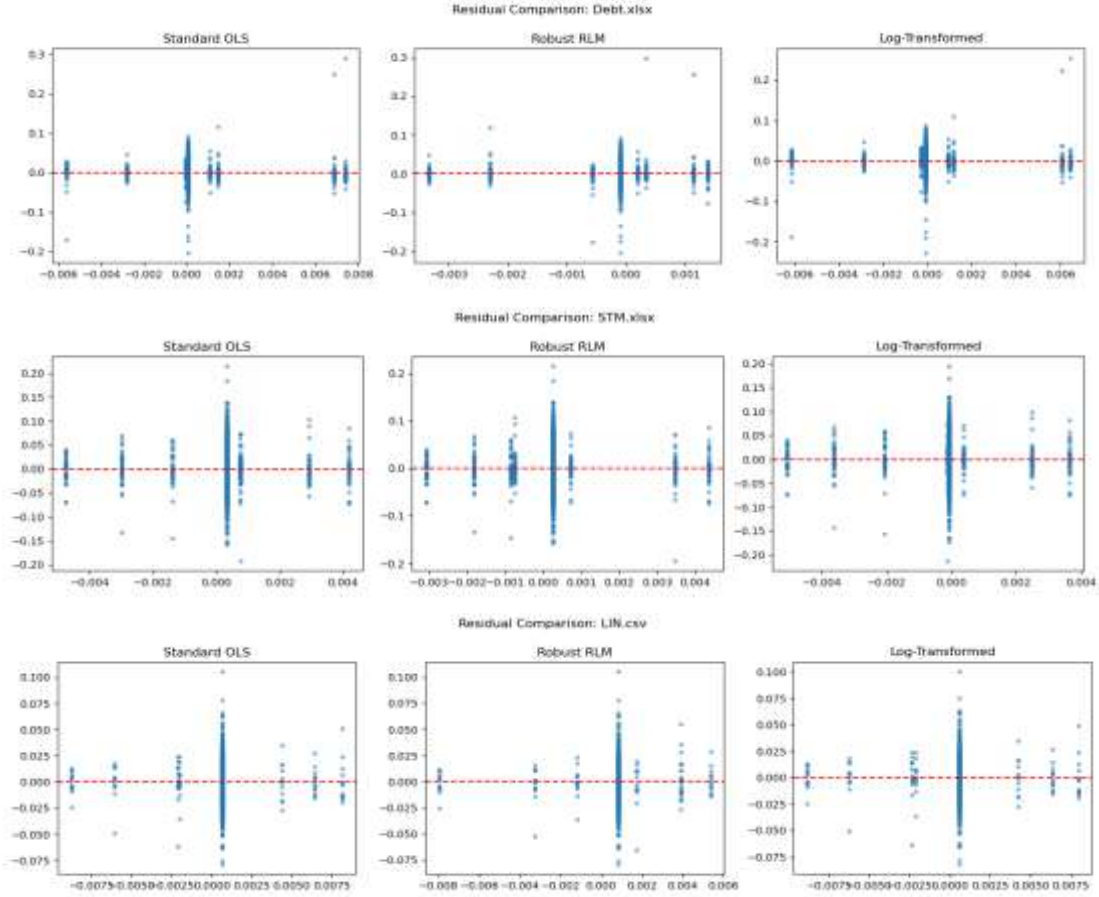


图 7 部分样本虚拟变量回归的敏感性分析

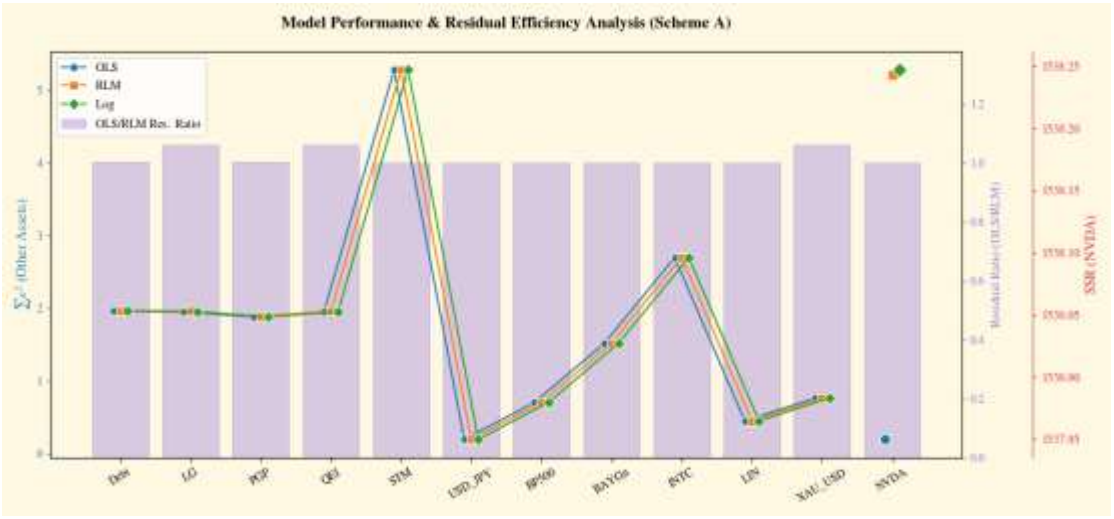


图 8 OLS&RLM&Log 变化折线图

随后我们同样对 4.4 引入的小波分析模型进行了敏感性分析，在变化标准化方法和对数据加入噪声分别进行拟合之后，不同的金融指数之间的结论仍然基本一致。

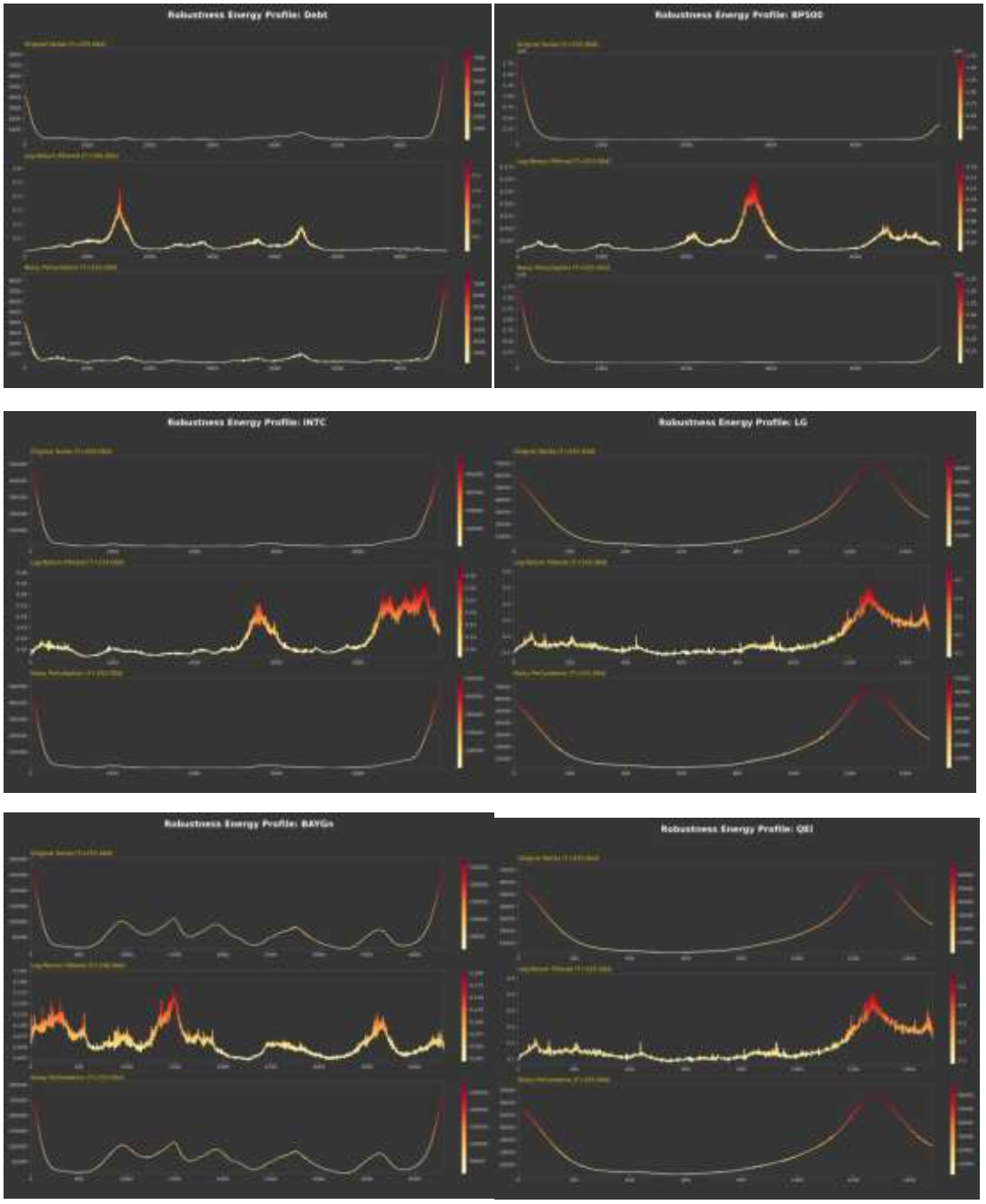


图 8 针对金融指数小波分析的敏感性检验

4. 幸福指数指标体系及权重确立

4.1 构建指标体系

我们基于 2020 年到 2026 年的数据建立指标体系并进行研究。普通人所感知的“幸运”，在很大程度上源于对所处社会环境“安全性”与“可控性”的潜意识评估。因此，我们将指标体系分为以下 5 个大类：

公共卫生事件：流行病的防控措施分布范围广且具有不可抗性，不仅直接影响健康，更通过限制社交和经济活动改变了幸福感的基准线。在此我们构造虚拟变量，以 JHU 公布的数据库的结束时间为基准，在此之前设置为 1，表示疫情管理发生，其余为 0。
经济波动：侵蚀个人对未来的掌控感，这种“不可控性”会催生无力感，并常被归因为“坏运气”。 在此我们用美国 30 年债券以及黄金的收益率为代表数据进行研究。
社会舆情：社会舆情作为集体情绪的实时指标，通过社会认同与情绪感染机制，内化为个人对时运的判断（Michael & Mark，2018）。当集体情绪低落时，个人也容易受到影响。在此我们用 Google 搜索指数中关于“TIRED”词条的声量来进行分析。
气候：直接影响日常情绪，研究表明特定温度（如 13.9°C）与更高的幸福感相关（Osaka University，2012）。在此我们以香港地区每天的平均湿度为代表。
社会安定状况：直接威胁生命与财产安全，动摇人类最基础的安全需求（马斯洛《人类动机理论》，1943），其恶化会动摇“幸运”感知的根基。由于犯罪率很难统计到高频数据，因此我们以百度指数中关于犯罪的头条内容来表示社会安全性的波动。

随后通过对所选指数的$Kendall\ \tau$相关性分析，我们发现这些指标之间，无论是线性还是非线性的相关性都较弱，数据之间的异质性特征明显。进一步证明指标选取的范围较广，具有代表性和广泛性。故保留进行后续进一步的论证研究。

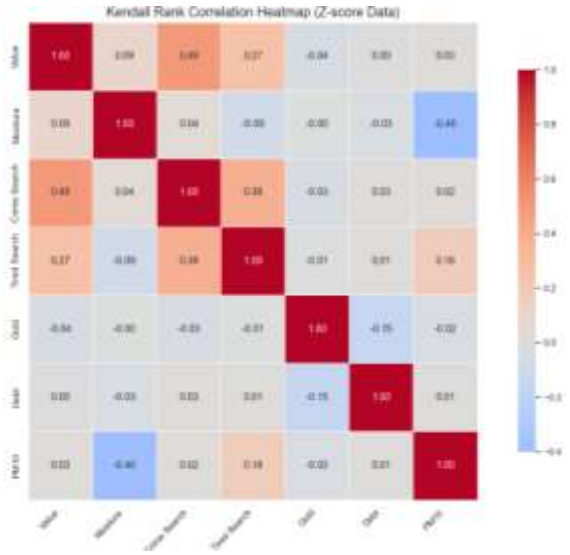


图 9 指标体系的异质性分析

4.2 熵权法

熵权法是一种基于信息论中熵的概念而产生的客观赋权方法。它通过分析每个指标所包含的信息量来确定指标的权重。其方法相对客观，在过往研究中曾大量被拿来进行经济社会指标的权重计算，且数据已经相对异质，无需考虑主成分分析等方法。因此我们可以直接采用熵权法计算权重，随后将结果与现实基础进行对比验证是否合理。

(1) 计算指标比重

$$p_{ij} = \frac{y_{ij}}{\sum_{i=1}^{m} y_{ij}} \tag{16}$$

(2) 计算信息熵$\boldsymbol{e_j}$

$$e_j = -k \sum_{i=1}^{m} p_{ij} \ln(p_{ij}) \tag{17}$$

$k = \frac{1}{\ln(m)}$. e_j 反映了指标 j 的信息分散程度.

(3) 变异系数 $\boldsymbol{d_j}$

$$d_j = 1 - e_j \tag{18}$$

(4) 计算熵权 $\boldsymbol{w_j}$

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j} \tag{19}$$

4.3 权重结果

本研究采用熵权法对 7 个异质性指标进行客观赋权，并结合定性分析判断权重结果的合理性，旨在消除主观偏好对“个人幸福感”测度的干扰。根据表 3 所示的测算结果，权重分配呈现出显著的层级特征，深刻反映了 2020 年至 2026 年间宏观环境对个体福祉的冲击机制：

(1) **核心影响因子的识别：COVID-19 发生期 (W=18.78%)** 占据最高权重，其信息熵值（E=0.9922）最低，表明该指标在时间序列上具有最显著的信息差异度和辨识力，是决定该时期公众幸福感波动的首要内生变量。作为样本区间内（2020-2026）最大的外部冲击，该结论基本合理，具有现实依据。

(2) **生存环境与心理负荷的共振：PM10 (16.92%)** 与 **疲惫搜索指数 (15.42%)** 紧随其后。这种权重分布揭示了外部物理环境压力与内在心理耗竭感之间的耦合关系，证实了环境质量与精神卫生指标在度量个体“幸运度”或“福祉感”时的高贡献率。

(3) **经济波动与社会安全感的影响路径：黄金收益率 (13.33%)** 与 **美债收益率 (11.41%)** 的权重分布，体现了避险资产波动对公众心理预期的间接映射；而 **犯罪搜索指数 (14.80%)** 的权重则量化了社会安全感知对个体幸福感的边际影响。

整体而言，差异系数 (D) 与最终权重 (W) 的分布逻辑自洽，权重结果与现实社会观察高度契合，为后续构建“科学避险规则”提供了具备数据支撑的优先级序列。

表 3 熵权法权重结果

指标	信息熵 (E)	差异系数 (D)	最终权重 (W)
Covid-19 发生期	0.9922	0.0078	18.78%
PM10	0.9930	0.0070	16.92%
疲惫搜索指数	0.9936	0.0064	15.42%
犯罪搜索指数	0.9939	0.0061	14.80%
黄金收益率	0.9945	0.0055	13.33%
美债收益率	0.9953	0.0047	11.41%
空气湿度	0.9961	0.0039	9.35%

5. 模型拓展: 幸福指数是否受到“黑色星期五”或者是其他日期规律的影响？

5.1 黑色星期五以及日期独立分解研究

通过构建虚拟变量线性回归模型，对日期属性的解释效力进行了量化评估。结果显示，核心解释变量——代表“13 号”与“星期五”重合的复合虚拟变量，回归模型中表现出**统计学上的非显著性**。

5.2 小波分析结果

随后我们进行小波分析后同样发现，依照频率能量图以及输出的频率特征序列我们得到，在研究的所有天数（2192 天）中，被模型识别为“好日子”的有 277 天，而“坏日子”则为 398 天。其中好日子和坏日子的周期中位数分别为 334 和 360。整体的占优周期近似为 365 天。为了保证模型结果的可靠性，我们进行了 COI 分析，结果显示在 95%的情况下“坏日子”统计显著。因此从数据的角度上来看，尽管从显著性的角度上来看，的确在特定的日期幸福指数相较于其他时候更高或者耕地。但由于周期的不稳定性，并且从结果来看“好日子”和“坏日子”的显现频率与标准公历年份下的频率设定并不相同。因此从整体上来看，尽管真的存在这样的一个规律，但是对公众的体验上来看，“好日子”和“坏日子”的日期位置是一直往下变动的，所以并不存在一个固定的日期。**因此严格上来讲并不存在“好日子”或者“坏日子”。**

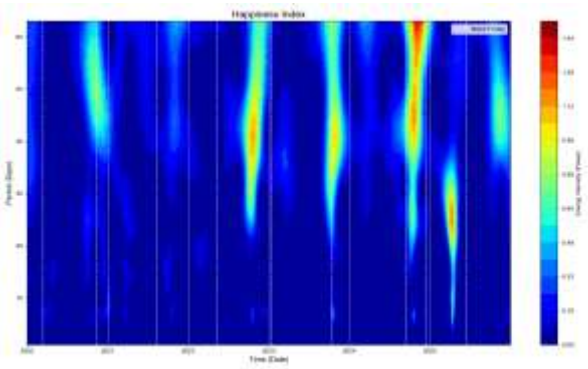
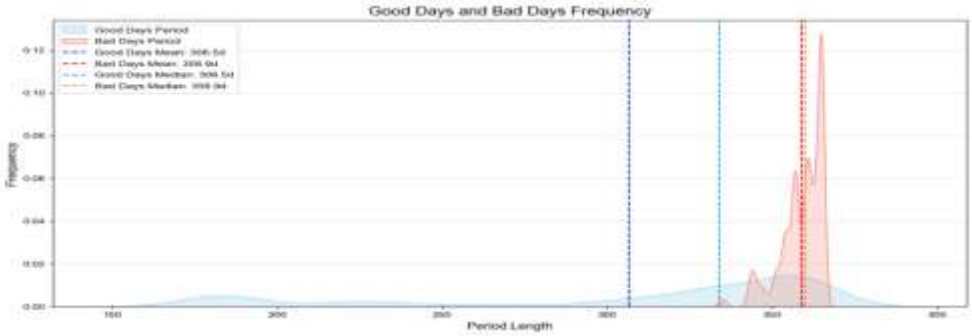


图 10 小波分析能量热力图



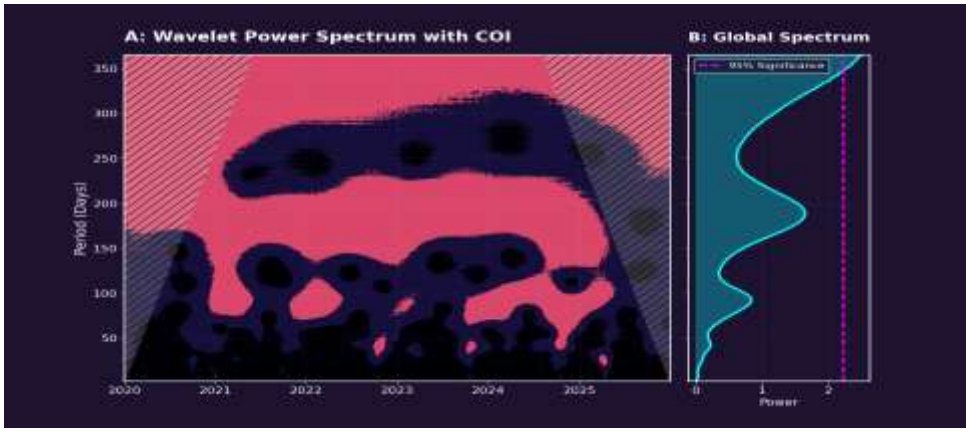


图 11 幸福指数小波分析子图

6. 基于幸福指数的“好日子”和“坏日子”预测

6.1 模型设定

在这里我们将前面小波分析得到的频率能量特征序列$D = \{(x_i, y_i)\}_{i=1}^{n}$输入到XGBOOST 模型之中，利用梯度下降的方法，最小化回归误差与惩罚函数的目标函数，进行特征值的筛选和预测值的更新。

(1) 初始化预测

$$\hat{y}_0 = argmin_c \sum_{i}^{n} l(y_i, c) \tag{20}$$

(2) 计算梯度 g_i、Hessian h_i
(3) 用 Gain 选择最优分裂，长出一棵树结构 q
(4) 计算每个叶子权重 $w_j^{\backslash *}$
(5) 更新预测 $\hat{y}^{(t)} = \hat{y}^{(t-1)} + \eta f_t$
(6) 输出模型并预测 $y^{(x)} = \sum_{t=1}^{T} f_t(x)$

6.2 模型结果

XGBOOST 模型的重点不在于得到模型的参数，而是通过实时同步数据更新下一次的预测结果。从结果上来看，利用 XGBOOST 进行预测的精度较高，**预测值与真实值的差基本控制在[0,1]的范围之间**。说明模型拟合效果较好。因此若想要对好日子或坏日子进行预测，可以考虑先经过小波分析得到频率能量特征序列，之后再将序列输入到 XGBOOST 模型中让模型经由梯度下降函数和惩罚函数逐步迭代致最优。

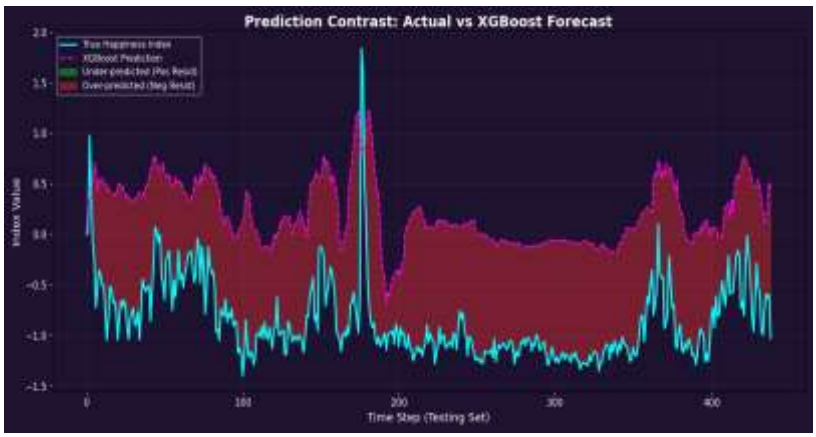


图 12 XGBOOST 模型预测结果

6.3 敏感性分析

在进行 Monte-Carlo 分析以及 What-if 分析之后，得到处理后的图像与未经过噪声处理的 XGBOOST 模型基本一致，进一步证明 XGBOOST 模型在预测和确认“日期迷信”频率上的优良性。

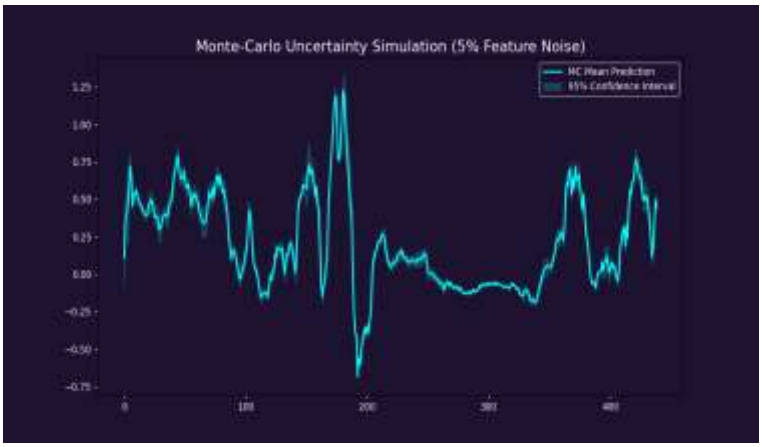


图 13 XGBOOST 模型预测的敏感性分析

7. 结论

本研究从在西方社会长久以来的迷信“黑色星期五”为起始点，首先探讨了黑色星期五对“负面事件”的发生频率是否有影响，之后通过对气候、经济、社会舆情、社会安定以及公共卫生事件等多源异质数据的量化分析，系统性地探究了“黑色星期五”是一种事实还是迷信，以及是否存在一种日期效应与公众的幸福感知有关。结果发现：

(1) **迷信证伪**

回归分析结果显示，在所有的金融指数中，代表“黑色星期五”的虚拟变量的系数统计不显著，其系数对所研究的金融市场 12 种指数的边际贡献也微乎其微。这表明“黑色星期五”的负面效应在金融市场并不存在。并且由于金融市场的指数收益率本身就为一种极其敏感的指标，在西方社会中也与公众息息相关。因此，基于上述结论以及相关的文献综述，我们基本可以认为“黑色星期五”仅为一种文化心理投射，而非事实。

(2) **周期效应体感缺失**

对于“负面事件”和公众幸福感知指数是否存在“好日子”或者是“坏日子”这一问题。在经过小波分析，观察和输出频率能量序列以及全局母序列之后，我们发现对于周期而言，能够依据频率判定为“好日子”和“坏日子”的周期都不相同，且这些周期的中位数都均不等于任何人们可以判定为规律的数字（如 7，30，365 或者农历日历时间），因此这种效应也不存在。

模型优势

- **多源异质数据的融合力。**我们针对题目的要求，将幸福指数的指标体系指标之间类型差异化，成功将金融收益、环境指标（PM10）与社会情绪搜索指数等因素整合，打破了单一类别数据的局限性。
- **方法的科学严谨性**。我们采用虚拟变量对日期效应进行“解构”。这种方法能够剥离出单纯的“星期五”因素和“13 号”因素，从而精准捕捉其交叉项的净效应。然后我们使用**熵权法**，利用数据本身的信息熵来决定权重，并将权重结果与现实依据进行对比，确保其合理性。这样我们既确保了“幸运指数”评估体系的客观性，又符合“科学依据”的要求。
- **极强的稳健性**：在每一个任务中，我们都使用多重的敏感性分析方法（如，$OLS, RLM, Log-transformed$ 对比），进行 Monte-Carlo 模拟，证明结论在不同算法下基本保持一致，证明了模型方法的稳健性。

模型局限性

- **滞后效应考量不足**：我们回归模型主要侧重于当日效应，当考虑重大负面事件（如 Covid-19）对公众心理产生的滞后性影响时，往往会由于特征太多、模型自由度的缺失而导致结论不明显。
- **文化差异的简化**：本模型虽基于全球数据，但对不同文化背景下“幸运”定义的异质性处理仍有优化空间。未来可引入地理加权回归（GWR）来分析特定地区的文化敏感度

参考文献

[1] Moutidis, Iraklis, and Hywel T. P. Williams. “Good and Bad Events: Combining Network-based Event Detection With Sentiment Analysis.” Social Network Analysis and Mining, vol. 10, no. 1, Aug. 2020

[2] Fidrmuc, Jan, and J. D. Tena. “Friday the 13th: The Empirics of Bad Luck.” Kyklos, vol. 68, no. 3, July 2015, pp. 317–34.

[3] T J Scanlon, R N Luben, F L Scanlon, N Singleton. "Is Friday the 13th bad for your health?" British Medical Journal 1993;307:1584, 18 December 1993, https://doi.org/10.1136/bmj.307.6919.1584.

[4] Vicki M. Bier, Statistical methods for the use of accident precursor data in estimating the frequency of rare events, Reliability Engineering & System Safety, vol. 41, no. 3, 1993, pp. 267-280, ISSN 0951-8320, https://doi.org/10.1016/0951-8320(93)90079-E.

[5] Petersen, Trond. “The Statistical Analysis of Event Histories.” Sociological Methods & Research, vol. 19, no. 3, Feb. 1991, pp. 270–323. https://doi.org/10.1177/0049124191019003002.

[6] Maslow, A. H. “A Theory of Human Motivation.” Psychological Review, vol. 50, no. 4, July 1943, pp. 370–96.

[7] Tsutsui Yoshiro. Weather and Individual Happiness. 1 Mar. 2012, ir.library.osaka-u.ac.jp/repo/ouka/all/14034/?lang=1&mode=1&opkey=R173991538920290&idx=10.

[8] Michael A Hogg, Mark J Rinella, Social identities and shared realities, Current Opinion in Psychology, Vol. 23, 2018, Pages 6-10, ISSN 2352-250X, https://doi.org/10.1016/j.copsyc.2017.10.003.